DOCUMENT RESUME

ED 362 210 IR 016 366

AUTHOR Tucker, Susan A.

TITLE Evaluating Interactive Instructional Technologies: A Cognitive Model.

PUB DATE Jan 93

NOTE 37p.; In: Proceedings of Selected Research and Development Presentations at the Convention of the Association for Educational Communications and Technology Sponsored by the Research and Theory Division (15th, New Orleans, Louisiana, January 13-17, 1993); see IR 016 300.

PUB TYPE Information Analyses (070) -- Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS Cognitive Psychology; Data Collection; Decision Making; *Educational Technology; *Evaluation Methods; *Feedback; *Models; Objectives; *Program Improvement; Synthesis; Values

IDENTIFIERS *Cognitive Models; Interactive Decision Making; *Interactive Teaching

ABSTRACT

Strengths and weaknesses of prevailing evaluation models are analyzed, with attention to the role of feedback in each paradigm. A framework is then presented for analyzing issues faced by evaluators of interactive instructional technologies. The current practice of evaluation relies heavily on 3 models developed over 20 years ago: (1) the objectives-based model of R. W. Tyler (1942); (2) the decision-making approaches exemplified by the models of M. Provus (1971) and D. L. Stufflebeam (1983); and (3) values-based approaches such as those of R. E. Stake (1967, 1982) and M. Scriven (1967, 1972). To varying degrees, these three approaches lack methods for systematizing the feedback of both the evaluated and the evaluators. An interactive evaluation model is suggested that attempts a synthesis of the strengths of the other models informed by the research base of recent developments in cognitive psychology. The model consists of four phases. The first, in which feedback figures strongly, is negotiating a paradigm to focus the evaluation and specify major questions, sources of evidence, and standards. The second is collecting and analyzing data. A third phase is judging alternatives and synthesizing a delivery matrix of recommendations. The final phase is helping the client process decision and choice mechanisms to improve the program. Three figures and one table illustrate the discussion. (Contains 164 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED 362 210

Title:

Evaluating Interactive Instructional Technologies: A Cognitive Model

Author:

Susan A. Tucker

IR016364

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Michael Simonson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Introduction

This paper has three main purposes. The first purpose is to analyze the strengths and weaknesses of three prevailing evaluation models, with special attention to the role of feedback (overt or covert) in each paradigm. The second purpose is to present a framework for analyzing issues faced by evaluators of interactive instructional technologies. Finally, the implications of applying this framework to macro-level evaluation of interactive instructional technologies will discussed.

Questions, Questions, Questions

Professionals in the business of affecting human performance cannot escape evaluation. When making program decisions, a systematic process of judging must be used. Compound the pressure of instructional decision-making with ever-changing interactive technology delivery systems, and the inherent problems surrounding evaluation expand into an exponential migraine.

With these problems come a wave of questions. Where do you start? Who are your real clients? How do you decide upon the major questions to structure your study? How do you address hidden agendas? How sufficient are objectives-based evaluations when working with innovative programs? How do you balance quantitative and qualitative data collection methodologies? Sure, it's cheaper than hiring outsiders--but does formative internal evaluation really work? How do you decide the evaluation is complete? Most importantly, after you have conducted this study, how do you feedback your hard-earned evaluation results to critical audiences? Call these professionals what you will-- trainers, instructional designers, course developers, instructional technologists, curriculum planners, teachers, program evaluators--they have not yet reached consensus about how to answer these questions.

Since the late 1960s, authors have addressed the need for systematic program evaluation models. Undoubtedly, the use of models to order educational processes facilitates the conception of problems and the perception of interrelationships within these problem contexts. While each model may view the same set of interrelationships, it is inevitable that certain concerns of the model builders will differ due to personal frames of reference (Alkin, 1968). Dillon (1981) extends this personal dimension of modelling to include "prefabricated configurations of concepts" or "patterns of coherence that we expect to find."

The modeling of social systems involves another potential danger. Education's use of analog models drawn from the "hard" sciences is troublesome to the extent that empirical models represent only "observable" realities. The human dimension in instruction requires models which represent both observable and subjective realities. This dilemma of inherent change is further compounded in evaluation when the evaluator threatens to question existing structures, make covert values visible, and test commonly accepted myths (Tucker & Dempsey, 1991).

In 1971, Daniel Stufflebeam raised several issues of concern to evaluators. These issues have yet to be addressed adequately by prevailing evaluation models. He notes four problems in the application of experimental designs to evaluation. First, selection of experimental designs conflicts with the principle that evaluation should facilitate continuous program improvement. According to a colleague of Stufflebeam, "experimental design prevents rather than promotes changes in the treatments because ... treatments cannot be altered if the data about differences between treatments are to be unequivocal" (Guba, 1969, p. 34). Second, traditional research methodologies are useless for making decisions during the planning and implementation of a project. Rather, the stress on controlling operational variables creates a contrived situation and blocks the collection of natural and dynamic information. Third, the typical research design does not apply to the "septic" conditions of most evaluation contexts. The evaluator is not interested in establishing highly controlled conditions within which

universally true knowledge can be generalized. Instead, "one wishes to set up conditions of invited interference from all factors that might ever influence a learning (or whatever) transaction" (Guba, 1969, p. 33). As a final point, internal validity is gained at the expense of external validity. Clearly, equating evaluation models with empirical models is limiting to dynamic program development concerned with credibility, (Lincoln & Guba, 1985), particularly in largely undocumented areas like instructional technology programs.

Diversity of major evaluation models

The current practice of evaluation relies heavily on three models developed over twenty years ago: (1) Tyler's (1942) objectives-based model, (2) the decision-making approaches exemplified by Provus' discrepancy model (1969, 1971) and Stufflebeam's (1983) CIPP model, and (3) values-based approaches such as Stake's responsive schema (1967, 1982) and Scriven's goal-free model (1967, 1972).

Model diversity emerges from the various emphases placed on each of these tasks. Another source of variance is how the evaluator relates to the client system. For example, Pace's (1968) analysis of evaluation models indicates that large, complex, and of longitudinal evaluations may require a different model -- one that considers a broad range of social and educational consequences. Holistic judgments require assessing more than program objectives. Instead, expand the evaluation to question other program dimensions such as expected and unexpected consequences, cultural characteristics of the setting, and the processes of program delivery. Evidence gathered to answer these questions should be both quantitative and qualitative, including the systematic collection of personal perceptions.

There are strengths and weaknesses in each approach which have emerged over the ensuing years. Rather than focus on their well-publicized strengths, the rest of this section will summarize some inconsistencies or areas which have not been addressed by each model. This critique is motivated by the need to understand why evaluation of interactive instructional technology projects has been fraught with difficulties and is often a dissatisfying experience for both the evaluator and the evaluated.

To demonstrate the different emphases placed in various evaluation concerns, Figure 1 compares three prevailing evaluation approaches: Tyler's instructional objectives (IO) model, Stufflebeam's decision- making (DM) approach, and the values-based (VB) approaches of Scriven and Stake. Special attention is given to the following criteria: a systematic methodology which utilizes a combination of quantitative and qualitative data sources to portray a more holistic reality; helpfulness toward program improvement; the evaluator's openness to make values overt and to facilitate the planning of probable futures for a program. The rest of this section analyzes the strengths and weaknesses of each model in more depth.

Insert Figure 1 about here

---

1. Instructional Objectives Approaches

Tyler's model has several merits. It provides valid, reliable, and objective data for an evaluation. It allows the evaluator to indicate attained and unattained objectives. On the other hand, strict application of the Tyler Model creates difficulties. It ascertains student outcomes but ignores the contexts and processes that lead to these final outcomes. The statement of objectives in behavioral terms is a long and often tedious procedure which limits sources of evidence to purely quantifiable data. In essence, it is too easy to avoid questions about the worth of the objectives themselves. This is especially true if the evaluation is conducted after the objectives

are formulated. When this happens, the evaluation is often limited to making superficial revisions of performance objectives.

In response to these difficulties, more flexible "neo-Tylerian" models have emerged (Taba & Sawin, 1962; Cronbach, 1963; AAAS Commission on Science Education, 1965). Tyler's data sources have been expanded to include observations on teaching method, patterns of classroom interaction, physical facilities, and student motivation. "Neo-Tylerian" ideas have contributed to the field's shift from a terminal focus to one which synthesizes both process and product elements. The emphasis of evaluation is on facilitating instructional improvement. Two major limitations persist. One, the evaluators do not incorporate client feedback into the proposed evaluation. Second, futures planning is neglected.

2. Decision Making Approaches

A second group of evaluation theorists, notably Provus and Stufflebeam, believe the evaluator's task is one of delineating information to be collected, planning a data collection methodology, and helping decision makers use the information collected. It is the responsibility of the decision maker, not the evaluator, to make judgments from this information. Four questions are basic to this approach:

- What should the objectives be?
- How should the program be designed to meet these objectives?
- Is the program design being carried out effectively and efficiently?
- Was the program worth what was invested in it considering the products achieved? (Reasby, 1973, p. 23).

While advancing the field by specifically addressing standards, Provus (1971) identified several weaknesses in his model. He concludes that the major weakness with the discrepancy model seems to stem from failure to understand the limits on institutional behavior set by field conditions, the intricacies of the decision-making process, and the origin and use of criteria in that process. Context evaluation is not addressed and the exhaustive use of behavioral standards may limit the creative, adaptive responsiveness of a program staff. Provus's approach does not clearly identify the type and ultimate use of information to be collected at each stage. The model could also be faulted for being incapable of evaluating rapid, large-scale changes characteristic of early instructional technology projects. Decision-makers are not always rational, yet the model assumes such behavior. Finally, the Discrepancy Model does not address how evaluators are recruited, trained, and maintained in the system.

The Phi Delta Kappan Committee on Evaluation, chaired by Daniel Stufflebeam (1971), has perhaps exerted more influence than any group in attempting to lead the educational profession away from excessive reliance on classical research models. Instead, an evaluation model is offered which assists decision makers in pinpointing their values so that they can best be served by the decisions made. This model, known as CIPP, specifies four types of evaluation: Context, Input, Process, and Product (Stufflebeam, 1983).

Though mechanically total, the CIPP model excludes overt values in its schema. According to Stufflebeam, the evaluator's role is one of collection, organization, and analysis of "relevant" data for "relevant" decision makers. Overt judgment of the intrinsic worth of the program's objectives is not considered by the evaluator or the clients. The entire question of values is kept tacitly in the decision-makers' domain. Another limitation of the CIPP model is its inability to answer two basic questions. How do evaluators and/or clients know which standards are operant? Further, what processes are necessary to enable the decision maker to apply value criteria?

3. Values-Based Approaches

Two major examples of values-based approaches to program evaluation are considered in this paper: Scriven's Goal-Free Model and Stake's responsive schema.

Scriven's (1967; 1972; 1978) Goal-Free model is in definite contrast to those evaluators advocating objectives-based or decision-making approaches. According to Scriven, the distinction between the roles and goals of evaluation are often intentionally blurred. Whatever its role, the goals of evaluation are always the same -- to estimate the merit, worth, or value of the thing being evaluated. Scriven goes on to point out that the subversion of goals to roles is very often a misguided attempt to allay the anxiety of those being evaluated. The consequence of this type of distorted evaluation could be much more undesirable than the anxieties evoked. As an alternative, Scriven (1972) declared that both summative and formative evaluations will increasingly become "goal-free." Scriven stresses the evaluation of actual effects against a profile of demonstrated needs. One of the many roles of the evaluator is to examine the goals of the educational program and judge the worth or value of these goals against some standard of merit. From the data he has collected, the evaluator can determine whether these objectives are being met.

One of the great strengths of this model is addressing the role and values of the evaluator. In spite of the model's strengths, several questions still remain answered. How can a client insure that external evaluators properly judge actual effects of a program, whether planned or not? What standards are there to judge whether a goal-free evaluator is not arbitrary, inept, or unscrupulous in his actions? How does one judge how well a goal-free evaluator has interpreted the "demonstrated needs" of a project?

The theme throughout many of Stake's writings (1967, 1970, 1972, 1973, 1975, 1982) is that an evaluator must do his best to reflect the nature of the program and not focus on what is most easily measured. Furthermore, he writes less about the link between evaluation and decision making than other evaluation theoreticians. Like Scriven, Stake believes that both descriptions and judgments are essential and basic acts of any evaluation and should be combined to portray an educational program. He goes on to describe the role of evaluator as a technician who can provide both relative and absolute judgments. In this respect, Stake takes a more conservative stance than Scriven. Stake (1973) recommends that the evaluator should not take an absolutist position regarding the program's goals because this is likely to make clients less willing to cooperate in the evaluation.

The Stake model exceeds most models in attempting to describe and judge the entire educational enterprise, rather than examining outcomes alone. In terms of process and scope, the Stake model deals more comprehensively with a number of evaluation needs. The assessment of evaluation procedures in the descriptive matrix is unclear as it appears that procedures would be described with transactions. However, procedures are selected prior to the learning experience. An effective method for analyzing the role of values in the program can be found in this model as can a methodology for discovering which values are being served. On the critical side, Stake's approach does not appear to provide for evaluating decision alternatives during the structuring of the learning experience. Furthermore, he does not provide for systematic feedback in program development. While the instructional program is looked at in terms of antecedents, transactions, and outcomes, underlying the whole approach is a perspective that looks at evaluation from the terminus.

Polemics in Program Evaluation

Evaluation is ever present and continues to operate, both overtly and covertly. Furthermore, evaluation is only as good as the environment that the evaluator is able to coalesce and share among all significant players. Approaching this cooperative state can be facilitated by Polanyi's (1962, 1966, 1975, 1978) constructs of collecting

multiple perceptions in order to approach a shared reality. He also advocates trying to make tacit perceptions overt rather than being unconsciously controlled by these perceptions, and including extended dwelling in an evaluation to experience more than the surface and first impression phenomena. In essence, out of these diverse perceptions we can identify critical polemics as signposts which expand but do not fix our capability to describe and judge. That is, we can use these polemics to avoid a closed view of evaluation and open our perspective to the variety of factors which can influence evaluator's and client's judgments. Figure 2 summarizes some of these major polemics and their continua.

Insert Figure 2 about here

The Need for a Different Model to Evaluate Instructional Technologies

As summarized in Figure 2, the three prevailing approaches possess similarities and differences. Special attention is directed to the presence or absence of several criteria: (1) using a systematic methodology which combines quantitative and qualitative data sources to portray a more holistic reality; (2) helping toward program improvement; (3) providing evaluator's feedback to make values overt; and (4) facilitating program planning regarding probable futures. To varying degrees, the instructional objectives approach, the decision-making approach, and the values-based approaches all lack methods for systematizing the feedback of both the evaluated and evaluators.

Given the aforementioned variance in evaluation models as well as diversity in clientele and program purposes, it is apparent that judgments about program worth can vary accordingly. It is the contention of this paper that the perceptions of both the evaluated and the evaluators need to be made as overt as possible. This cognitive process will enable the acts of informed judgment making and program improvement. Support for this position has been persuasively advanced by Weckworth:

> "First, there is no one way to do evaluation; second, there is no generic logical structure which will assure a unique 'right' method of choice. Third, evaluation ultimately becomes judgment and will remain so, so long as there is no ultimate criterion for monotonic ordering of priorities; and fourth, the critical element in evaluation is simply: who has the right, i.e., the power, the influence, or the authority to decide." (1969, p. 48).

It is the purpose of this section to offer an interactive evaluation model (Kunkel & Tucker, 1983; Tucker & Dempsey, 1991) which addresses some of the weaknesses of the aforementioned models and strives for credible synthesis of their strengths. This is especially difficult to achieve in contexts reflecting diverse client and audience values such as instructional technology programs. A macro-level evaluation model is therefore perceived as a body of interrelated criteria. These criteria help determine major questions, sources of evidence, and standards for judging the findings and making recommendations. Finally, it is informed by the research base of recent developments in cognitive psychology. Figure 3 summarizes the cognitive complexity of this process as a series of:

- contextually relevant exchanges between information inputs and feedback messages between the evaluation and clients;
- perceptual representations of this data; and
- resultant actions, responses and/or judgments regarding the data.

Insert Figure 3 about here

By making overt the vulnerable judgments of both evaluators and those being evaluated, a synthetic picture of what the evaluation intends to focus on can be negotiated. Through cycles of feedback between the evaluator and clients and negotiation using the feedback information, operative values can be embodied in the major evaluation questions asked. In addition, these values can serve as judgment criteria for decision-making. On the other hand, if the operative values are not known, distortion results in terms of what is questioned, how evidence is gathered, and how judgments are made.

In the model of evaluation proposed, certain value criteria are non-negotiable in the sense that along with accepting the evaluator personally, the primary audience must be made aware of and accept five criteria. That is, the model embodies holism, negotiation, evaluator vulnerability, multiple data sources, and is improvement-oriented and futuristic.

A Cognitivistic Model of Program Evaluation

The proposed model of evaluation relies greatly on developments in cognitive psychology. Cognitivism helps us understand the processes behind two evaluation truisms: "Information is power only in the hands of those who know how to use it" and "Evaluation is only as good as the information it is based upon." It is helpful to view evaluation as a complex activity which involves sequences of: gathering information within situations or contexts, representing this information in idiosyncratic perceptions, and then using this data for some sort of judgment, decision, or action.

Evaluators and decision makers are required to integrate information from many sources when forming an initial judgment, reviewing recommendations, and ultimately making a choice among alternative strategies. The quality of the decision hinges upon how well the rater is able to encode, store, retrieve, and integrate information from these multiple sources into a judgment. Unfortunately, this process is not easy. A piece of information must pass through many filters before it is recorded or retrieved. More than informational, feedback can also serve motivational functions (Kopelman, 1986). It can provide information about the correctness, accuracy and adequacy of the program. It encourages developing, revising, and refining task strategies, thereby laying the foundation for program improvement (Earley, 1988). Not surprisingly, decision makers use only a subset of available information in evaluating a program. In fact, information can delay and complicate decision making for many. When confronted with a problem, the decision maker brings his limited information and limited cognitive strategies and arrives at the best solution possible under the circumstances. Simon (1978) calls this "satisficing."

As described in the next section, research suggests decision-making is both a rational and irrational process. Decision-makers rely upon simplifying strategies or heuristics when judging complex environments. While these strategies can be helpful in avoiding cognitive overload, if applied inappropriately, heuristics can provide a significant source of bias and error in decision outcomes if applied inappropriately. To counter this potential bias, an evaluation model (see Figure 3) is proposed that:

- is interactive and negotiated between the clients, evaluators and participants;
- is based on the perceptions and values of many participants to be capture a holistic rather than positivistic reality;
- involves significant players throughout all four phases, not just at the beginning and the end; and
- systematically uses feedback to serve informational and motivational functions.

Phase 1: Generating Questions and Setting Standards

As depicted in Figure 3, Phase I has two levels of feedback loops. These loops consist of perceptual exchanges between the clients, evaluators and relevant audiences. These exchanges involve the perception of the task demand of the client and major questions synthesized by the evaluator to focus the evaluation.

Consider the first feedback loop. Here, the client introduces the task demand to the evaluator. The client's presentation can be influenced by both overt and covert factors, how stable the client's organizational environment appears to be, prior history of experiences with evaluation, and general belief structures. It becomes readily apparent that evaluation serves many functions besides error reduction. For example, it can provide evidence of the client's competence, signal impending action, and can protect the client politically.

Next, the evaluator perceives the task. This perceptual representation is shaped by a variety of factors. These factors include the evaluator's prior work experiences, his ability to decode semantic and semiotic messages, the perceived priority or strength of demand of the task, recognition of political forces, confidence, and stakeholder boundaries.

The first output of the evaluator is an attempt to identify the program's operational questions and standards expressed by the client. This could be done at any phase of the instructional process (e.g., analysis, design, development, implementation, or dissemination). This completes the first cycle of information inputs.

If an evaluator uses only the first feedback loop to develop his questions and set standards, the risk of perceptual misunderstandings between the evaluator and client are still present. Recognizing this risk, experienced evaluators often confirm their identification of the major questions by seeking feedback from the client. Clients are asked how relevant these questions are to their interests and needs (versus a more research-oriented perspective which would give the evaluator sole authority to determining what is valuable and relevant). Factors which can influence the client's feedback message include: the client's expectations about short and long term program outcomes, the personal significance of the questions, penalty costs of whether or not to gather the information, the client's perception of the evaluator's degree of expertise, and rapport established between the client and evaluator.

When feedback is readily available, the risk of serious initial judgment errors is not as high because feedback provides information upon which to base corrective adjustments. Consider Hogarth's (1981) likening this process to aiming at a target. The evaluator may engage in a series of exchanges with the client. The intent of these exchanges is to focus the evaluator's perception of the task or "target." Then the evaluator forms a mental representation of the perception. For example, in the first phase of evaluation, this representation serves to eliminate, substitute, or strengthen the evaluator's perceptions about the questions being asked. For example, perceptions of any hidden agendas emerging from the first cycle of information inputs are considered. Given the prior experience of the evaluator, the task may be encoded as objectives-based, management-based, or values-based, or a composite of all three.

Finally, this cycle ends with the evaluator negotiating an evaluation plan. This plan or paradigm serves as a blueprint to guide the ensuing evaluation. Three components make up this paradigm: questions, evidence, and standards. Questions to guide the study are framed, potential sources of evidence clarified, and standards or qualities by which to judge the findings are made overt. Negotiation of the paradigm allows for the systematic participation of all relevant stakeholders. This negotiation should be done during the first phase of the evaluation, thereby reducing some of the perceptual errors inherent in models such as Scriven's goal-free approach. It has been the experience of the author that negotiation serves as the first step in co-opting even

the most negative stakeholders. This is accomplished by being open to their critical questions and emerging perceptions.

Implications of Phase 1

Helping stakeholders generate questions is affected by one's capacity and motivation to generate alternative hypotheses for a given question. (For more thorough discussions of hypothesis generation refer to the work of Higgins & King, 1981; Eco and Sebeok, 1983; Fisher, Gettys, Maming, Mehle, & Boca, 1983; Mayseless & Kruglanski, 1987.) For example, in situations where time pressures are great, a client or evaluator may be tempted to focus solely on addressing pre-defined objectives. In settings where formative evaluation provides useful feedback for product development and revision, its design is often constrained by available resources such as time, money, personnel, and facilities.

Extending the example of formative evaluation to interactive instructional technology further, it appears that two major questions are posed. One deals with the content and technical quality of the product and the other deals with its learnability within authentic contexts. More specific questions generated by these two foci could deal with product initiation, design, content development, media development, production, testing and maintenance (Foshay, 1984). Geis (1987) suggests that feedback is needed regarding the content and sequencing of instructional events as well as the nature of learner control of the content or message as sent. Consider the case of instructional hypermedia. Table 1 summarizes just some of the questions that could emerge when one is open to more holistic views of evaluation, hopefully more accurately reflecting the medium's true nature.

Insert Table 1 about here

Besides these two major questions, each stakeholder appears to have idiosyncratic questions of interest. Rather than reaching consensus on the questions, an effort is made to solicit the complete realm of questions and then synthesize them into broad questions which allow multiple perspectives. Fifteen years of evaluation field narratives suggest this is a viable strategy and management research seems to support this as well. The research of Walsh and associates (1988) argues that increased coverage and decreased consensus are important early in the decision making process to help define the problem or task. Once the group comes to understand the information environment, however, a consensus around a narrower view (i.e., evaluation paradigm negotiated during feedback loop 2) is beneficial. The ability to read a decision environment and capture the major evaluation questions and the belief structures behind these questions is the essence of shared veridicality.

Questions are also influenced by an individual's capacity to see the values or standards that are behind the questions. Why was the question asked in the first place? Standards describe the results that are desired or should be achieved upon satisfaction of a task. To facilitate performance, researchers contend that standards need to be specific as well as identify qualities and quantities of output expected (Baird, 1986). Specific standards serve as motivational devices for individual performance (Locke et al., 1981; Taylor, 1983) and can anchor future decisions of the client (Huber & Neale, 1986). This phenomenon may explain the evolution of "building block" evaluations which focus on lower level, fragmented tasks which are more easily identified and documented.

Instructional designers lack consistent formative evaluation guidelines regarding products that aim at interactive, integrative skill development. It is difficult (but not impossible) to evaluate situations which allow learners to practice multiple skills. Compounding this complexity is the presence of unpredictable

navigational paths and a need to assess variable performance standards. In fact, interactive instructional design seems fraught with many aversive prior learning experiences. For starters, many managers expect cost overruns and schedule slippage. Another fear involves the losses attributed to the new product's possible failure spreading to established products. Finally, while many designers recognize that prevailing linear and iterative strategies result in "piecemeal" products, they lack alternative design methods peculiar to this technology.

Program and product performance is compared to a norm or standard. Usually, the norm is the maximum achieved by an optimal allocation policy. Invariably, performance is found wanting (Busemeyer, Swenson, & Lazarte, 1986). Optimal policies cannot be specified without perfect knowledge of the goal state. For many real decisions, however, only imperfect and vague information about the objective and solution is available. Even if specifications exist, management may not have made it clear whether the intent is to meet specifications or exceed them. Hence, there is a perceived need for evaluation. This seems to be particularly true of interactive technologies such as instructional hypermedia where the operant criteria have yet to be defined.

And what happens when the product's instructional and business goals are not clearly communicated to the technicians and instructional designers? What often results is loyalty to functional standards versus the total plan because immediate functional rewards upstage measures of organizational performance. For example, in university environments where product development is almost exclusively driven by external funding, the product can have high priority within the grant but low priority throughout the rest of the institution.

Standards are often very difficult to calculate, and it seems unreasonable to expect that the typical client knows the optimal solution a priori. For example, developers' and users' judgment of product quality extends beyond hardware and software. Quality involves technical, educational, and financial attributes; installation issues; low maintenance; high reliability; compatibility with other equipment already in place; ease of upgrading; and multi-platform access. These are standards that are typically not incorporated into questions asked by developers. However, it is reasonable to expect that program managers and designers can improve their standards and rules for optimal policies iftraining incorporates informative feedback (Busemeyer, Swenson & Lazarte, 1986). And the corollary is that while this rational explanation may be possible, the evaluators must be prepared for instances where the client intends the evaluation to serve functions other than an error reducing role. For example, clients may want the evaluation to give evidence of a key player's competence, signal impending action, or "cover their tracks."

Expectations about future outcomes strongly affect one's decisions (Abelson & Levi, 1985; Feather, 1982). Developers often devote much time examining the implications of alternative designs and the adequacy of the likely reward in view of the risks incurred. Because of the importance of such expectations, much research has been devoted to understanding how individuals estimate probabilities of future events. Evidence has accumulated that individuals use a general cognitive strategy, the "availability heuristic" (Tversky & Kahneman, 1973). In using this heuristic, individuals estimate the probability of events by the ease with which they can recall or cognitively construct relevant instances. Specifically, questions can serve as a valuable heuristic or cognitively simplifying strategy for participants in the evaluation process.

Not surprisingly, our field experience indicates that decision makers use only a subset of available information in evaluating a program. Rather, managers seem to rely on heuristics to simplify complex environments. While serving to avoid information overload, inappropriately applied heuristics can provide a significant

source of bias and error for decision outcomes. For instance, people seem inclined to overweigh certainty in their decisions (Kahneman & Tversky, 1979). As evaluators, we should anticipate this heuristic when clients generate evaluation questions--and strive for questions that capture the task's complexity rather than satisfy predetermined decision-making. The systematic use of shared cognitive feedback and negotiation to "cleanse" the questions is seen as an effective strategy to minimize this bias.

Finally, framing a question can lead to a perceptual bias which may potentially be elicited by either task or context characteristics. Kahneman and Tversky (1979) suggest that decision makers treat the prospect of gains much differently than the prospect of losses. Often, whether the decision maker is evaluating the prospect of gains or losses is simply a matter of the way a question is presented or phrased. Thus, the way questions as well as findings are framed (in terms of losses or gains) can influence decision makers' risk propensity and thereby their decisions (Bazerman, Magliozzi & Neale, 1985; Huber & Neale, 1986; Neale & Bazerman, 1985).

Phase 2: Description of Data Sources and Analysis of Alternatives

The quality of the ultimate choices made hinges upon how well the clients and evaluators are able to attend to, encode, store, retrieve, and integrate information from multiple sources into a judgment. Ilgen, Fisher, and Taylor (1979) identified the information source as a particularly critical determinant of information utilization. They go on to warn that individual differences often influence information receptivity, credibility, and use.

Unfortunately, this process is not easy. A piece of information must pass through many filters before it is recorded or retrieved. Decision makers have a propensity to use only a subset of available information when judging a program's worth. This irregular (and often irrational) process may be more systematic if feedback loops are used. This loop or cycle consists of: (1) information inputs; (2) perceptual representations; and (3) outputs regarding the evidence gathered to answer the questions posed in phase one. Two feedback loops appear to operate during this second phase.

In the first feedback loop in this phase, the evaluator collects data. This process is affected by: the stability of the organizational environment; the length of the time the evaluation has committed to dwelling in the system; and the nature of the balance between qualitative and quantitative evidence. Acquiring information prior to a decision has two major risks. First, a tenuous evaluator could overacquire information and incur excessive costs. Second, an overconfident evaluator might underacquire information and incur excessive risks of decisional error. In any event, the costs of gathering additional information may be immediate and easy to estimate, but the benefits of doing so can be unclear and often long delayed (Connolly & Thorn, 1987).

Next, the evaluator analyzes the obtained data. This representation is affected by a myriad of factors such as: the evaluator's information load capacity; his audience analysis; the amount of anticipated versus unanticipated information revealed; audience and client analysis; and whether a compensatory or noncompensatory data synthesis model is being used (Einhorn & Hogarth, 1981; Billings & Marcus, 1983). Noncompensatory models minimize the questions, criticisms and divergent data. Compensatory models allow the representation of cognitively complex and sophisticated strategies for information integration.

After the data receives its initial analysis, the evaluator reports initial data findings to the client. This reporting procedure must consider many factors. Some crucial variables include: the amount of feedback processing time available to the client; the client's relative emphasis on program documentation versus program

improvement; and whether the evidence can be shared with the clientele all at one time or presented over a period of time.

Perceptual errors can result if the evaluator does not receive a message from the client regarding the accuracy of this initial data report. By scheduling a regular exchanges, the client's initial certitude about the original evaluation plan (as well as hidden agendas) can be ascertained. In the process, a sort of error checking occurs about the perceived accuracy of the data thus far collected as well as an early measuring of the evaluator's credibility. This is also the time to assess the client's tunnel vision and blind spots and selective perception.

The evaluator cognitively represents the client's messages in light of commitments (sometimes complementary and sometimes competing) between the client and evaluator. This representation takes the form of either improving or simply documenting the program. A choice point also involves the balance between qualitative and quantitative data. Obviously, this takes skill in balancing positive and negative information. This perceptual recycling of data perceptions is necessary in light of research showing human information acquisition is often weakly guided by normative, cost-minimizing considerations. This occurs even when serious efforts are made to simplify the task, provide high motivation and focus attention on balancing information costs and benefits. Clients consistently underpurchase "good" information and "bad" information is consistently overpurchased. One part of the explanation is client difficulty in discerning the validity of different information sources, even after repeated exposure. Another possible explanation is the certainty of the cost involved in acquiring information as against the uncertainty of its benefits in reducing decision errors. A risk adverse client will tend to overpurchase, a risk seeker to underpurchase.

Finally, the evaluator edits the initial data analysis and reports the revisions along with the implications of preliminary judgments given the evidence. Once again, influencing variables include the amount of time available to the evaluation for information feedback to the client, the kind of heuristics which accompany the data, the way information is "framed", and the form of information display.

Implications of Phase 2

Let's continue the example of formative evaluation of interactive technology products and/or projects. The actual process of data collection can generate as many questions as it answers. Just consider the following issues:

- How continuously should data collection be conducted: during rapid prototyping, draft form testing, final testing, or during each stage of the developmental process?
- What qualitative and quantitative methods should be used: self-evaluation, expert review or student review (Montague et. al., 1983); try-out and revision testing sessions (Geis, Burt, & Weston 1985); draft and revise and field-test (Dick & Carey, 1985); one-to-one testing; group testing; extended testing (Stolovitch, 1982; Komoski & Woodward, 1985); peer debriefing; triangulation; negative case analysis; and/or audit checks?
- Who should be included as data sources: learners (expert versus novice), developers, subject area experts, teachers and trainers, native informants, audience specialists, gatekeepers, sponsors, former learners, editors?
- How many people should be involved--what is an "adequate" size?
- What are desirable characteristics of the data sources: representative versus select (e.g., highly verbal or high aptitude), active versus passive, internal versus external to project, continuous versus one-time involvement?
- Where should data be collected: in-house versus field, using simplified versus progressively more complex and less familiar systems?

- When should data collection stop? For example, should evaluation continue until redundancy of learner comments occurs? When is the criterion performance level reached?

The client may accept the evidence "as is", request additional evidence, or even seek to reject the evaluation. For example, the use of user feedback in instructional materials development tends to be supported in research (Baker, 1974; Andrews & Goodson, 1980; Stolovich, 1982; Wager, 1983; Weston, 1986). In general, there does not seem to be an indication that one method of gathering feedback is superior to another. Rather, findings suggest that materials that have undergone formative evaluation with users were superior to original versions. While there is agreement that user feedback promotes product improvement, few clear guidelines exist regarding how to build this feedback systematically into the development process.

Knowledge of client expectations *and* behavioral characteristics are essential for evaluators when gauging the receptivity to their activities and ultimately to the information gathered. Available research in this area suggest that decision makers are:

- only weakly responsive to various normative considerations such as information quality and cost (Pitz, 1968; Wendt, 1969);
- substantially responsive to normatively irrelevant factors such as total information available (Levine, Samet & Brahlek, 1975);
- slow to show learning in repeated decision situations (Lanzeta & Kanaref, 1962; Wallsten, 1968; Busemeyer, 1985); and
- not consistent in their need for either overpurchasing or underpurchasing information (Pitz et. al., 1969; Hershman & Levine, 1970) though risk seekers show a tendency to underpurchase information.

A two tiered approach to data sharing is important for both cognitive and political reasons. The data is most likely to be heard and used by clients if they believe it is true and has utility. This satisfaction index seems to rise when client perceptions about data accuracy is sought before presenting data of a judgmental nature. Additionally, the form of information display can encourage or discourage certain methods of cognitive processing, given the propensity of humans to adopt strategies which minimize their effort. For example, organizing data into a table makes quantitative information easier to use, increasing its impact upon choices (Russo, 1977). Similarly, when data is described in words instead of numbers, decision-makers abandon strategies which use mental arithmetic (Huber, 1980). As Slovic (1966;1975) indicated, the more effort required, the more likely it is that clients will ignore or misuse information. Cognitive strain may cause decision-makers to resort to simplifying strategies or heuristics, many of which lead to biased responses such as preference reversals and violations of transivity. They take information as presented and change the strategy to suit the display rather than transform the data to fit the strategy. When evaluators recognize this process and the impact of framing, they can design evaluation data displays which passively encourage better information processing (Levin et.al., 1985).

Consider Snyder and Swann's (1978) proposal that in the testing of questions and hypotheses (about themselves or other people or events), people predominantly seek to confirm their pre-existing notions. This implies a pervasive insensitivity to disconfirming evidence, and (presumably) a reluctance to generate alternatives. The implications for how evaluators can share information with clients are profound.

Phase 3: Making Judgments

Many evaluators neglect to address their role in the act of judging, preferring to perceive themselves as untainted by values issues. This is counterintuitive. Judgment pervades the total process, from the selection of questions and data sources to the ultimate decisions reached and choices made. Phase 3 presents a series of informational inputs, perceptual representations, and actions in order to generate judgments about program worth.

After receiving the evaluator's final descriptions, data analyses, and tentative alternatives, the client reacts in the form of another feedback message. The nature of this feedback will be tempered by the client's perceptions of certitude about the findings and expected gains versus losses. Many researchers are pessimistic about managers' abilities to interpret information from complex environments accurately. Managers are thought to suffer from selective perception (Dearborn & Simon, 1958), strategic myopia (Lorsch, 1985), tunnel vision (Mason & Mitroff, 1981), and blind spots (Murray, 1978).

Given the client's reactions, the evaluator itemizes and begins to edit the pool of possible alternatives in preparation for generating judgments. Variables that seem to influence this editing process include the evaluator's capacity to visualize a continuum of concrete and abstract instances, the duration of events, and the client's self-efficacy status.

Editing results in the evaluator rating or valuing alternatives against the standards specified in Phase One. Besides the amount of processing time available for the evaluation, what is at stake is the quality of the alternatives envisioned as well as the heuristics used for conveying these alternatives to the client (for example, availability, adjustment, anchoring, and representativeness).

It appears that some decision strategies are used more than others. The issue is frequency of use rather than effectiveness of stategy. Recent research suggests expectation models over-intellectualize the cognitive processes people use when choosing alternative actions (Schwab, Olian-Gottlieb, & Heneman, 1979; Fischhoff, Goitein, & Shapira, 1983; Mitchell & Beach, 1990). For example, formal analytic strategies are seldom used even by people who know about them. Isenberg (1984) adds that even when they use these logical strategies they seldom accept the results that run counter to their intuition. In certain cases, intuition turns out to be more accurate than analytic strategies (Hammond, Hamm, Grassia, & Pearson, 1987). Mitchell and Beach (1990, p.3) capture this concept well with the following statement: "Formal analytic strategies require a great deal of concentration, time, effort, and skill and very few decisions are worth all that. Moreover, for even very important decisions the formal strategies often seem too coldly intellectual".

Observations of professional decision makers engaged with on-the-job decisions suggests that few decisions involve explicit balancing of costs and benefits, let alone the use of behavioristic probability models (Peters, 1979). Mintzberg (1975) reported that most decisions involved only one option rather than many options. The decision was whether to go with that option rather than a choice among competing options.

Simultaneously, the evaluator and the client develop judgments. Their ability to form sound information-based judgments appears to be grounded upon many factors, including:

- ability to reason on three levels: deductively, inductively and abductively;
- ability to manipulate overt and covert information normatively and heuristically;
- ability to negotiate or at least share perceptions;
- clarity of expectations, both internal and external; and
- amount of cognitive dissonance and uncertainty of the decision to be reached.

Out of the judgments represented in the prior step, the evaluator synthesizes a delivery strategy and accompanying recommendations. Besides dealing with the standard variables of determining the timing, written and/or oral formats, and the framing of the delivery matrix, the evaluator needs to weigh the impact of priming. Priming is a process through which particular information becomes activated or more readily accessible to recall (Wyer & Srull, 1980). Priming has been found to impact the type of action plan developed by an individual. In fact, those individuals who were faced with pressures imposed by a difficult goal relied more heavily on the plan they had been provided than those with more general goals (Earley & Perry, 1987). Furthermore, recommendations are more likely to be maintained if the outcomes are positive and more likely to be changed if the outcomes are posed negatively.

Implications of Phase 3

Given the fact that decision makers are required to integrate information from numerous sources when forming a judgment, the quality of the resultant judgments relies on the value each stakeholder places on this information. For example, Foshay (1984) suggests three criteria of interest to training vendors:

- cost-effectiveness (such as reliability and validity of measurement, return on investment, corporate reputation, and contractual expectations for project);
- compatibility with project management system in terms of yielding timely, useful project status information and impacting employee morale and productivity;
- compatibility with the design system such as generating information specific to design and development decisions.

While the research base is limited, our experience suggests that clients judge situations as instances of types with which they are familiar (pattern recognition). They tend to respond best where their own prior successful solutions are presented as the recommended strategies. This can be a problem when trying to enhance an already effective operation with a novel strategy. In addition, there seems to be a difference in whether the "client" making the judgments is an individual or a collective. Fischhoff et. al. (1983) suggest that while individual clients may adapt to recommendations and judgments that will further their gains more easily than their losses, a group of clients is influenced by social consensus. Team members who have been recently successful are more responsive to the needs of others and are less likely to make independently serving judgments.

Self-esteem seems to be a related variable as well. Recent research (Weiss & Knight, 1980; Knight & Nadel, 1986) suggests low esteem people gather more information about possible solutions before implementation and perform better on tasks where the one best solution must be identified (i.e., optimizing). High self-esteem people seem to search for less information before trying a solution, performing better on tasks with obvious solutions, strict time constraints, or where information search is costly. Knight and Nadel (1986) suggest that high self-esteem managers would be less likely than managers low in self-esteem to experiment with new policies or solutions when confronted with negative performance feedback.

While Fischhoff's findings bode well for cohesive teams, many interactive development projects consist of several (sometimes competing) teams. This specialization can give rise to functional managers who begin to insist on making all decisions pertaining to their respective stages and inputs from other teams are not welcomed. The longer the development cycle, the greater the possibility of communication rifts. A compounding problem concerns accountability and receptivity to evaluation feedback in competing environments. If functional managers know they will be held accountable for anything that goes wrong, there is a tendency to hold up work on

the new product pending written notice (Foshay, 1984). This sign-off process can dramatically affect the openness to risk taking.

People's cognitive representation of judgmental tasks may conceive of "probabilities" as indices of belief intensity rather than as ratios of favorable chances (Cohen, 1982). They also might conceive of events as nonrandom (Cohen, 1982) and fail to take into account all the potentially relevant information or all the potentially relevant alternatives (Hintikka & Hintikka, 1983). It appears that the motivation to respond to judgmental questions may be affected by at least two needs: the need for cognitive structure and the fear of invalidity. Preliminary research by Mayseless and Kruglanski (1987) suggests that revision of judgments might be slower under a high fear of invalidity but faster under a high need for structure. "In short, fear of invalidity might induce a tendency to be conservative, whereas need for structure might induce a tendency to be excessive" (p. 180). For example, in novel, open-ended environments such as instructional hypermedia projects, designers may often experience an acute need for guiding structure to which they might respond by adopting inappropriately high levels of self-assurance and a closed-mindedness to alternative points of view.

This heightened need for cognitive structure is assumed to promote an early closure to a solution or "cognitive freezing" (Freund, Kruglanski & Schpitzajzen, 1985). For evaluators, this suggests the many risks inherent in how early to give client feedback as well as the nature of the information's significance. Get to know your client cognitively!

But there are risks for evaluators as well. For example, high structure evaluators who rely on checklists for interactive instructional materials revision risk neglecting some critical components since little prescriptive documentation exists at this juncture. It seems more defensible to involve a diverse set of leaders and opinion makers as reviewers if the goal is to trend acceptance of materials. There is still little to guide us in deciding how to choose experts (as well as evaluators) and how to guide their task or structure their output (Geis, 1987).

A person with a high need for cognitive structure is likely to inhibit the generation of competing alternatives to a given recommendation as such alternatives might appear to threaten his existing schema. Previous research manipulating the need for structure in such ways found that individuals in which this need was aroused tended to base their judgments more on (1) early information, rejecting subsequent data, thereby exhibiting "primacy effects", and (2) on pre-existing stereotypes, in this sense being theory rather than data driven. Furthermore, high versus low "need for structure" individuals tended more to seek comparison with similarly minded others likely to support their views and opinions (Mayseless & Kruglanski, 1987). High "need for structure" individuals were characterized by: higher initial confidence in their judgment; more confidence in early information provided by the search; fewer requests for information; and high final confidence in their judgment.

Functionally opposite to the need for structure is the fear of invalidity. This motivation has to do with the desire to avoid mistakes in judgment in terms of their perceived costliness. Where high need for structure promotes a freezing of the judgmental process, fear of invalidity may often promote unfreezing. That is, an increased tendency to respond positively to alternative solutions to the existing situation and/or an increased sensitivity to information inconsistent with the prevailing order and negative feedback. Previous research manipulating the fear of invalidity found that high fear of invalidity individuals suppressed the magnitude of primacy effects. Instead, they had a tendency to translate stereotypes into discriminatory judgments (Kruglanski & Freund, 1983). High fear of invalidity individuals had lower initial confidence, less confidence in early information, requested more information, and had higher final confidence. Overall, fear of invalidity might induce a tendency to be conservative.

Priming constraints are not to be ignored (Tversky & Kahneman, 1981). For example, when presenting negative feedback, about the effectiveness of interactive technology in a training program, individuals who are provided with negatively framed recommendations (i.e., taking an action to prevent losses) are more open to questioning the status quo and consider risky alternative strategies while those provided with positively framed recommendations (judgments to protect gains) are more likely to choose safer, more predictable outcomes. Individuals who have experienced recent losses are more likely to accept informative feedback and are more open to risky alternatives than usual (Fischhoff & Beyth Marom, 1983; Fischhoff, Goitein, & Shapira, 1983).

Finally, little research has been conducted concerning the effects of standards on client and evaluator judgment. It seems plausible that standards might anchor rater judgments by providing a natural starting point against which to evaluate performance outcomes. It also seems logical that performance-based priorities (weighting) would simplify the rating process (Naylor, Ilgen, & Pritchard, 1980) but would exacerbate anchoring effects for low priority standards. Consequently, the validity of performance ratings would be highest when standards are specific and highly weighted (Neale, Huber & Northcraft, 1987). Results of the Neale et. al. study suggest that performance standards do not influence raters' performance related judgments. As they suggest, it would be instructive to test whether evaluators and clients are capable of incorporating differential weighting into their appraisal judgments.

Finally, the recent work on image theory (Beach & Mitchell, 1987; Mitchell & Beach, 1990) seems very promising, particularly the compatibility test. The notion is that intuitive, automatic decision making (and even some deliberative decision making) relies on a simple comparison of each alternative of interest with a limited set of relevant standards called images. Images serve as informational representations and consist of: (1) value images such as the decision makers' ethics, imperatives, and vision of what is appropriate; (2) trajectory images such as future aspirations, goals, and agendas; and (3) strategic images such as plans, tactics and forecasts used to attain the desired goals. If a decision alternative is incompatible with these images, it is rejected. If it is not incompatible, it is accepted as the final decision or as one of the alternatives that is passed on to a more complex mechanism that selects the best among them. The concept of image seems to be related to cognitivists "schemata" and "scripts" (Graesser, Woll, Kowalski, & Smith, 1980; Anderson, 1983) and control theory's "template" (Lord & Hanges, 1987).

Phase 4: Arriving at Decisions and Choices

The final phase of this evaluation model deals with decision and choice mechanisms. Decisions consist of presenting clients with alternatives packaged in a certain form, such as a set of outcomes and probabilities. After searching a perceptual representation of options, the decision maker edits and evaluates these alternatives. Alternatives are compared with each other by rationally calculating the degree of preference or using intuitive tests of compatibility and profitability. Choosing the "best" alternative involves strategies such as elimination by aspects (Tversky, 1972) and prospect theory (Kahneman & Tversky, 1979) and profitability test (Mitchell & Beach, 1990). In addition, affective variables such as mood and the context of the alternatives of a choice can be important in making decisions. As discussed earlier, we know that people tend to avoid risk when alternatives are gains and seek risks when alternatives are losses.

The context in which decisions occur appears to give them meaning. In addition, past successess and failures in similar contexts seems to provide guidance about what to do about the current decision. If the current decision is virtually identical to a past decision stored in memory, it is considered to be recognized and

automatic (Beach, 1964). Framing results when the contexts of prior similar decision memories are used to go beyond the information that is presented by the current situation alone and interpret new contexts. Thus, we might expect that manipulating the framing process could have a significant effect on decision making and choosing a final behavior (Rachlin, 1989). In novel contexts where the exceptional is encountered, it seems that the process becomes much less automatic.

The decision mechanism is influenced by whether the client is a novice decision-maker or expert, whether the decision environment is self-initiated by the client or imposed from above or pressured from below and the degree to which the client is risk-taking or risk-seeking. The selection among decision strategies is often seen as a tradeoff between the amount of cognitive resources (effort) required to use each strategy and the ability of each strategy to produce an "accurate" effect (Beach & Mitchell, 1978; Johnson & Payne, 1985; Russo & Dosher, 1983).

The choice mechanism is an outgrowth of the decision mechanism. Its efficiency is contingent upon: the client's perceived self-efficacy; the environmental demands of the client's situation; the attractiveness of the alternatives and incentives for change; past expenditures of effort; short and long term performance valences; and the type of delivery strategy required by the decision.

Finally, the outcome and implementation of the ultimate choice results. As can be seen by the prior sequences, decision and choice are complex, cognitive operations of which evaluators must be more cognizant if genuine implementation of evaluation findings and recommendations is desired.

Implications of Phase 4

Two major types of decision strategies described in the literature are compensatory and noncompensatory models. Compensatory models represent cognitively complex and sophisticated strategies for information integration (Einhorn & Hogarth, 1981) which are indicated by the absence of the interactive use of cues (Billings & Marcus, 1983). Noncompensatory models are indicated by the interactive use of information cues in which a low score on one dimension cannot be compensated by a high score on another dimension (Billings & Marcus, 1983).

Compensatory strategies refer to either the linear model or the additive difference model. The linear model assumes that each dimension for a decision alternative is given a value for each alternative. Comparisons among alternatives are then based on these overall values and the alternative with the greatest value is selected. The additive difference model implies that decision makers compare alternatives on each dimension and then summing differences between dimensions. The summation of differences results in a preference for one decision alternative (Olshavsky, 1979). With both linear and additive difference models, a high value on one dimension "compensates" or counteracts low value on another dimension for the same decision alternative. Noncompensatory strategies involve the use of simplifying rules to reduce to complexity of the decision. The major noncompensatory models identified by Payne (1976) and others include conjunctive, disjunctive, lexicographic, and elimination by aspects strategies.

Frequently, clients must make choices under less than ideal conditions. Uncertainty results when people have incomplete information about the task and doubt is typically generated in a crisis situation when time is restricted for decision-making and there are unanticipated choice points have been presented. Increasing uncertainty is often associated with a decentralization of an organization's communication structure (Tushman, 1979) while increasing the threat often leads to a centralization of structure (Staw, Sandelands & Dutton, 1981). Consider the wide variety of organizations where interactive instructional technology is being created, from an individual's home to

large scale multinational corporations and how the nature of this environment can significantly affect the receptivity to make choices revealed by an evaluation.

A related factor appears to be the complexity of the task. Decentralized structures were found to be more efficient and resulted in fewer decision errors for complex and uncertain tasks (Shaw, 1981). Faucheaux & Mackenzie (1966) found that groups performing simple tasks evolved toward a centralized structure while those performing complex tasks did not. Recent researchers have generally supported these findings but caution that the relationship between uncertainty and structure is a complicated one that depends on additional factors such as the quality of information a client receives from individuals versus groups, insiders and outsiders like contracted evaluators, the skill of its leaders, and the particular sources of uncertainty (Argote, 1982; Fry & Slocum, 1984; Schoonhoven, 1981; Argote, Devades, & Melone, 1990).

This tendency toward centralization can be dysfunctional given centralized structures perform more poorly than decentralized structures for complex and uncertain tasks (Snadowsky, 1972). This appears to occur because members at the hub of centralized networks experience overload under high uncertainty conditions and centralized structures are very vulnerable to this increased overload. As more group members perceive that the information needed to reach a decision and make a choice resides throughout the group rather than one member, we expect decentralized structures to emerge. While this projection is fascinating to contemplate, it is anticipated that it will occur with much discomfort and resistance.

Four basic mediator strategies which have implications for evaluators: press, compensate, integrate and inaction (Carnevale and Conlon, 1988). Press refers to efforts to reduce disputant stakeholders' aspirations and occurs when evaluation mediators do not value client aspirations and they perceive that there is little common ground. Compensation deals with efforts to entice client disputants into agreement. For example, this occurs when evaluators value each stakeholder's aspirations, agreement appears likely, and there is little chance that integrating will be successful. Integration occurs when there are efforts to discover options that satisfy the disputants' aspirations or when mediating evaluators value parties' aspirations and perceive that there is common ground. Integration is used when there is a good chance of achieving a mutually acceptable solution. A final choice is inaction by which the mediating evaluation lets the parties handle the dispute on their own.

Summary

The systematic combination of evaluation and cognitive processes can guide the decision maker in a direction that continuously improves his or her performance. To varying degrees, the instructional objectives approach, the decision-making approach, and the values-based approaches all lack consistent methods for systematizing the perceptions of both the evaluated and evaluators. This paper has presented an alternative evaluation framework consisting of four sequential phases: (1) negotiating a paradigm to focus the evaluation and specify major questions, sources of evidence, and standards by which to judge the findings; (2) collecting and analyzing data sources and reporting the emerging implications of various alternatives; (3) judging the alternatives and synthesizing a delivery matrix of recommendations; (4) helping the client process decision and choice mechanisms instrumental in delivering an improved program.

References

AAAS Commission on Science Education. (1965). An Evaluation Model and Its Application. Publication 65/9. Washington, D.C.: American Association for the Advancement of Science.

Abelson, R.P., & Levi, A. (1985). Decision making and decision theory. In G. Lindzey & E. Aronson (Eds.), Handbook of Social Psychology (3rd ed., Vol. I, pp. 231-309). New York: Random House.

Alkin, M.C. (1968). Towards an Evaluation Model: A System Approach. CSE Working Paper No. 4. Los Angeles, CA: Center for the Study of Evaluation.

Allo, G. (1961). Research on Values at the Crossroads of Modern Civilizations. Beirut: Institute de Recherce et de Formation en Vue du Developpement (IFRED).

Anderson, C.A. (1983). Imagination and expectation: The effect of imaging behavioral scripts on personal intentions. Journal of Personality and Social Psychology, 45, 293-305.

Andrews, D.H., & Goodson, L.A. (1980). A comparative analysis of models of instructional design. Journal of Instructional Development, 3(4), 2-16.

Argote, L. (1982). Input uncertainty and organizational coordination in hospital emergency units. Administrative Science Quarterly, 27(3), 420-434.

Argote, L., Seabright, M.A., & Dyer, L. (1986). Individual versus group use of base-rate and individuating information. Organizational Behavior and Human Decision Processes, 38, 65-75.

Argote, L., Devades, R., & Melone, N. (1990). The base-rate fallacy: Contrasting processes and outcomes of group and individual judgment. Organizational Behavior and Human Decision Processes, 46, 296-310.

Ary, D., Jacobs, L., & Razavieh, A. (1972). Introduction to Research in Education. New York, NY: Holt, Rinehart and Winston.

Baird, J.R. (1986). Improving learning through enhanced metacognition: A classroom study. European Journal of Science Education, 8, 263-282.

Baker, E.L. (1974). Formative evaluation of instruction. In W.J. Popham (Ed.), Evaluation in Education (pp. 531-585). Berkeley, CA: McCutchan.

Bazerman, M.H., Magliozzi, T., & Neale, M.A. (1985). Integrative bargaining in a competitive market. Organizational Behavior and Human Decision Processes, 35(3), 294-313.

Beach, L.R. (1964). Recognition, assimilation, and identification of objects. Psychological Monographs, 78, 22-37.

Beach, L.R., & Mitchell, T.R. (1978). A contingency model for the selection of decision strategies. Academy of Management Review, 3, 439-449.

Beach, L.R., & Mitchell, T.R. (1987). Image theory: Principles, goals, andn plans in decision making. Acta Psychologica, 66, 201-220.

Bednar, A.K., Cunningham, D., Duffy, T.M., Perry, J.D., Theory into practice: how do we link. to appear in G. Anglin (Ed.), Instructional technology: past, present and future. Denver, CO: Libraries Unlimited.

Billings, R.S. & Marcus, S.A. (1983). Measures of compensatory and non-compensatory models of decision behavior: Process tracing versus policy capturing. Organizational Behavior and Human Performance, 31(3), 331-352.

Brehmer, B., & Hagafors, R. (1986). Use of experts in complex decision making: a paradigm for the study of staff work. Organizational Behavior and Human Decision Processes, 38, 181-195.

Busemeyer, J.R. (1985). Decision-making under uncertainty--a comparison of simple scalability, fixed sample, and sequential sampling models. Journal of Experimental Psychology of Learning, 11(3), 538-564.

Busemeyer, J.R., Swenson, K.N., & Lazarte, A. (1986). An adaptive approach to resource allocation. Organizational Behavior and Human Decision Processes, 38, 318-341.

Carroll, J. (1984). Analyzing decision behavior: The magician's audience. In T.S. Wallsten (Ed.), Cognitive processes in choice and decision behaviors. New Jersey: Erlbaum Associates.

Carnevale, P.J.D. & Conlon, D.E. (1988). Time pressure and strategic choice in mediation. Organizational Behavior, 42(1), 111-133.

Cohen, L.J. (1982). Are people programmed to commit fallacies? Further thoughts about the interpretation of experimental data on probability judgments. Journal for the Theory of Social Behavior, 12, 251-274.

Connolly, T. & Thorn, B.K. (1987). Predecisional information acquisition--effects of task variables on suboptimal search strategies. Organizational Behavior, 39(3), 397-416.

Cronbach, L.J. (1963). Course improvement through evaluation. Teacher College Record, 64, 672-683.

Cunningham, D.J. (1987). Outline of an education semiotic. American Journal of Semiotics, 5(2), 201-216.

Dearborn, D.C. & Simon, H.A. (1958). Selective perception: A note on the departmental identification of executives. Sociometry, 21, 140-144.

Dick, W., & Carey, L. (1985). The systematic design of instruction. (2nd ed.). Glenview, IL: Scott-Foresman.

Dillon, G. (1981). Constructing Texts: Elements of a Theory of Composition and Style. Bloomington, IN: Indiana University Press.

Dunbar, R. & Goldberg, W.H. (1978). Crisis development and strategic response in European corporations. In C.F. Smart and W.T. Stanbury (Eds.), Studies in Crisis Management. Toronto, ON: Butterworth.

Earley, P.C. (1988). Computer-generated performance feedback in the magazine subscription industry. Organizational Behavior and Human Decision Processes, 41(1), 50-64.

Earley, P.C., & Perry, (1987). Work plan priming. Organizational Behavior and Human Decision Processes, 39, -302.

Eco, V., & Sebeok, T.A. (Eds.).(1983). The sign of three. Bloomington: Indiana University Press.

Eco, V. (1983). Horns, hooves, insteps: Some hypotheses on the types of abduction. In V. Eco & T.A. Sebeok (Eds.), The sign of three. Bloomington: Indiana University Press.

Einhorn, H.J., & Hogarth, R.M. (1981). Behavioral decision theory: Processes of judgment and choice. Annual Review of Psychology, 32, 53-88.

Faucheaux, C., & MacKenzie, K. (1966). Task dependency on organizational centrality: Its behavioral consequences. Journal of Experimental Social Psychology, 2, 361-375.

Feather, N.T. (1982). Expectations and action: Expectancy-value models in psychology. Hillsdale, NJ: Erlbaum Associates.

Fischhoff, B., & Bar-Hillel, M. (1984). Diagnosticity and the base-rate effect. Memory & Cognition, 12, 402-410.

Fischhoff, B., & Beyth Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. Psychological Review, 90(3), 239-260.

Fischhoff, B., Goitein, B., & Shapira, Z. (1983). Subjective expected utility: A model of decision making. In R.W. Scholz (Ed.), Decision making under uncertainty. Amsterdam: North Holland.

Fischoff, B., & MacGregor, (1982). Judged lethality: How much people seem to know depends upon how they are asked. Risk Analysis, 13(4), 229-236.

Fisher, S.D., Cettys, C.F., Maming, C., Mehle, T., & Baca, S. (1983). Consistency checking in hypothesis generation. Organizational Behavior and Human Performance, 31(2), 233-254.

Foshay, W.R. (1984). QA and QC: A training vendor's view of the formative/summative distinction. Performance and Instruction, 23(10), 15-17.

Freire, P. (1970 a.). Pedagogy of the Oppressed. New York, NY: Herder and Herder.

Freire, P. (1970 b.). Cultural Action for Freedom. Harvard Educational Review, Monograph No.1.

Freund, T., Kruglanski, A.W., & Schpitzajzen, A. (1985). The freezing and unfreezing of impressional primacy: Effects of the need for structure and fear of invalidity. Personality and Social Psychology Bulletin, 11(4), 479-487.

Fry, L.W. & Slocum, J.W. (1984). Technology, structure, and workgroup effectiveness: A test of a contingency model. Academy of Management Journal, 27(2), 221-246.

Geis, G.L., Burt, C., & Weston, C.B. (1985). Instructional development: Developmental testing. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Geis, G.L. (1987). Formative Evaluation: developmental testing and expert review. Performance and Instruction, 26(4), 1-8.

Goulet, D.A. (1971). The study of values. Harvard Educational Review, 41(2), 205-227.

Graesser, A.C., Woll, S.B., Kowalski, D.J., & Smith, D.A. (1980). Memory for typical and atypical actions in scripted activities. Journal of Experimental Psychology, 6, 503-515.

Guba, E. G. (1969). The failure of educational evaluation. Educational Technology, 9(5), 29-38.

Hammond, K.R., Hamm, R.M., Grassia, J., & Pearson, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. IEEE Transactions on Systems, Man, and Cybernetics, SMC-17, 753-770.

Harris, K.L., & Carnevale, P. (1990). Chilling and hastening: The influence of third-party power and interests on negotiation. Organizational Behavior and Human Decision Processes, 47, 138-160.

Hershman, R.L. & Levine, J.R. (1970). Deviations from optimum information-purchase strategies in human decision-making. Organizational Behavior and Human Performance, 5(4), 313-329.

Higgins, E.T., & King, G. (1981). Accessibility of social constructs: Information processing consequences of individual and contextual variability. In N. Cantor & J. Kihlstrom (Eds.), Personality, cognition and social interaction. Hillsdale, NJ: Erlbaum Associates.

Hintikka, J., & Hintikka, M.B. (1983). Sherlock Holmes confronts modern logic: Toward a theory of information seeking through questioning. In V. Eco & T.A. Sebeok (Eds.), The sign of three. Bloomington, IN: Indiana University Press.

Hogarth, R.M. (1980). Judgment and choice. New York: Wiley.

Hogarth, R.M. (1981). Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. Psychological Bulletin, 90, 197-217.

Huber, O. (1980). The influence of some task variables on cognitive operations in an information procesing decision model. Acta Psychologica, 45(1-3), 187-196.

Huber, V.L., & Neale, M.A. (1986). Effects of cognitive heuristics and goals on negotiation performance and subsequent goal setting. Organizational Behavior and Human Decision Processes, 38(3), 342-365.

Ilgen, D.R., Fisher, C.D., & Taylor, M.S. (1979). Consequences of individual feedback on behavior in organizations. Journal of Applied Psychology, 64(4), 349-371.

Isenberg, D.J. (1981). Some effects of time pressure on vertical structures and decision-making accuracy in small groups. Organizational Behavior and Human Performance, 27(1), 119-134.

Isenberg, D.J. (1984). How senior managers think. Harvard Business Review, 62(6), 80-90.

Johnson, E.J., & Payne, J.W. (1985). Effort and accuracy in choice. Management Science, 31, 395-414.

Jolliffe, S.D. (1990). Formative evaluation of interactive training materials. In B.L. Flagg (Ed.), Formative evaluation for educational technologies. Hillsdale, NH: Erlbaum Associates. (pp. 99-112).

Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47, 263-291.

Kahneman, D., Slovic, P., & Tversky, A. (1982). Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press.

Khandwalla, P.N. (1978). Crisis responses of competing versus noncompeting organizations. In C.F. Smart and W.T. Stanbury (Eds.), Studies in Crisis Management. Toronto, ON: Butterworth.

Kleinmuntz, D.N., & Thomas, J.B. (1987). The value of action and inference in dynamic decision making. Organizational Behavior and Human Decision Processes, 39, 341-364.

Knight, P.A., & Nadel, J. (1986). Humility revisited: self-esteem, information search, and policy consistency. Organizational Behavior and Human Decision Processes, 38(2), 196-206.

Komoski, P.K. & Woodward, A. (1985). The continuing need for the learner verification and revision of textual material. In D.H. Jonassen (Ed.), The technology of text. Englewood Cliffs, NJ: Educational Technology.

Kopelman, R.E. (1986). Managing productivity in organizations: A practical, people-oriented perspective. New York: McGraw-Hill.

Kruglanski, A.W. (1980). Lay epistomo-logic: Process and contents: Another look at attribution theory. Psychological Review, 87, 70-87.

Kruglanski, A.W., & Freund, T. (1983). The freezing and unfreezing of lay inferences: Effects of impressional primacy, ethnic stereotyping, and numerical anchoring. Journal of Personality and Social Psychology, 46, 503-518.

Kunkel, R.C. & Tucker, S.A. (1983). Critical evaluation skills at a time of retrenchment. Paper presented at the annual meeting of the Association for Supervision and Curriculum Development, Houston, TX.

Lanzetta, J.T. & Kanareff, V. (1962). Information cost, amount of payoff and level of aspiration as determinants of information seeking in decision-making. Behavioral Science, 7, 459-473.

Levi-Strauss, C. (1969). The Raw and the Cooked. New York, NY: Harper and Row.

Levi-Strauss, C. (1976). Structural Anthropology, Vol II. New York, NY: Harper and Row.

Levi-Strauss, C. (1979). Myth and Meaning. New York, NY: Schocken Books.

Levin, J.Pl, Johnson, R.D., Russo, C.P., and Deldin, P.J. (1985). Framing effects in judgement tasks with varying amounts of information. Organizational Behavior and Human Decision Processes, 36, 362-377.

Levine, M. (1975). A cognitive theory of learning: Research on hypothesis testing. Hillsdale, NJ: Erlbaum Associates.

Lincoln, Y.S. & Guba, E.G. (1985). Naturalistic inquiry. Beverly Hills, CA: Sage.

Locke, E.A., Shaw, K.N., Saari, L.M., & Latham, G.P. (1981). Goal setting and task performance: 1969-1980. Psychological Bulletin, 90, 125-152.

Lord, R.G., & Hanges, P.J. (1987). A control system model of organizational motivation: Theoretical development and applied applications. Behavioral Science, 32, 161-178.

Lorsche, J.W. (1985). Strategic myopia: Culture as an invisible barrier to change. In R.H. Kilmann, M.J. Saxton, & R. Serpa (Eds.), Gaining control of the corporate culture. San Francisco, CA: Jossey-Bass.

Mason, R., & Mitroff, I. (1981). Challenging strategic planning assumptions. New York: Wiley.

Mayseless, O., & Kruglanski, A.W. (1987). What makes you so sure? Effects of epistemic motivations on judgmental confidence. Organizational Behavior and Human Decision Processes, 39, 162-183.

Milburn, T.W., Schuler, R.S. & Watman, K.H. (1983). Organizational crisis. 2. Strategies and responses. Human Relations, 36(12), 1161-1179.

Mintzberg, H. (1975). The manager's job: Folklore and fact. Harvard Business Review, 53(4), 49-61.

Mitchell, T.R., & Beach, L.R. (1990). "...Do I love thee? Let me Count..." Toward an understanding of intuitive and automatic decision making. Organizational Behavior and Human Decision Processes, 47, 1-20.

Montague, W.E., Ellis, J.A. & Wulfeck, W.H. (1983). Instructional quality inventory: A formative evaluation tool for instructional developers. Performance and Instruction Journal, 22(5), 11-13.

Morris, J.D., Smith, A.B. (1988). Using focus groups to evaluate instructional media: A case study. Educational Technology, 27-32.

Murray, E.A. (1978). Strategic choice as a negotiated outcome. Management Science, 24, 960-972.

Naylor, J.C., Ilgen, D.R., & Pritchard, R.D. (1980). A theory of behavior in organizations. New York: Academic Press.

Neale, M.A., & Bazerman, M.H. (1985). The effects of framing and negotiation overconfidence on bargaining behaviors and outcomes. Academy of Management Journal, 28, 34-49.

Neale, M.A., Huber, V.L., & Northcraft, G.B. (1987). The framing of negotiations: Contextual versus task frames. Organizational Behavior and Human Decision Processes, 39, 228-241.

Northcraft, G.B., & Earley, P.C. (1989). Technology, credibility, and feedback use. Organizational Behavior and Human Decision Processes, 44, 83-96.

Olshavsky, R.W. (1979). Task complexity and contingent processing in decision making: A replication and extension. Organizational Behavior and Human Performance, 24(3), 300-316.

Pace, R. C. (1968, December). Evaluation perspectives: 1968. Los Angeles, CA: UCLA Center for the Study of Evaluation. (ERIC ED 037828).

Payne, J.W. (1976). Task complexity and contingent processing in decision making: AN information search and protocol analysis. Organizational Behavior and Human Performance, 16, 366-387.

Peters, T. (1979). Leadership: Sad facts and silver linings. Harvard Business Review, 57(6), 164-172.

Pitz, G.F. (1968). Information seeking when available information is limited. Journal of Experimental Psychology, 76, 25-34.

Pitz, G.F., Reinhold, H., & Geller, E.S. (1969). Strategies of information seeking in deferred decision-making. Organizational Behavior and Human Performance, 4, 1-19.

Podsakoff, P.M., & Farh, J. (1989). Effects of feedback sign and credibility on goal setting and task performance. Organizational Behavior and Human Decision Processes, 44, 45-67.

Polanyi, M. (1962) Personal Knowledge. Chicago, IL: University of Chicago Press.

Polanyi, M. (1966). The Tacit Dimension. London, UK: Pergamon Press.

Polanyi, M. (1975). Meaning. Chicago, IL: University of Chicago Press.

Polanyi, M. (1978). Phenomenology of Perception. Tr.C.Smith. London, UK: Routledge and Kegan Paul.

Porsecanski, T. (1989). Ideologies and development: a report from South America. The American Journal of Semiotics, 6(2/3), 291-298.

Provus, M. (1969). Evaluation on ongoing programs in the public school system. In R.W. Tyler. (Ed.) Educational Evaluation: New Roles, New Means. Chicago, IL: N.S.S.E.

Provus, M. (1971). Discrepancy Evaluation. Berkeley, CA: McCutchan Publishing.

Rachlin, H. (1989). Judgment, decision and choice. New York: W.H. Freeman.

Reasby, H. (1973). Application of multiple evaluation strategies to comprehensive and innovative middle school programs. Unpublished Doctoral Dissertation, University of Washington.

Russo, J.E. (1977). The value of unit price information. Journal of Marketing Research, 14(May), 193-201.

Russo, J.E. (1978). Eye fixations can save the world: A critical comparison between eye fixations and other information processing methodologies. In H. Keith Hunt (Ed.), Advances in Consumer Research. Chicago, IL: Association for Consumer Research, volume 5, 561-570.

Russo, J.E., & Dosher, B.A. (1983). Strategies for multiattribute binary choice. Journal of Experimental Psychology: Learning, Memory & Cognition, 9, 676-696.

Schoonhoven, C.B. (1981). Problems with contingency theory-: Testing assumptions hidden within the language of contingency "theory". Administrative Science Quarterly, 26(3), 349-377.

Schwab, D.P., Olian-Gottlieb, J.D., & Heneman, H.G. (1979). Between-subjects expectancy theory research: A statistical view of studies predicting effort and performance. Psychological Bulletin, 86, 139-147.

Scriven, M. (1967). The methodology of evaluation. In B.O. Smith. (Ed.), AERA Monograph Series on Curriculum Evaluation. Chicago, IL: Rand McNally.

Scriven, M. (1972). Prose and cons about goal-free evaluation. Evaluation Comment, 3(4), 1-8.

Scriven, M. (1978). Goal-free evaluation in practice. Paper presented at the fourth annual meeting of the American Educational Research Association, Toronto, Ontario.

Sharp, G.L., Cutler, B.L., & Penrod, S.D. (1988). Performance feedback improves the resolution of confidence judgments. Organizational Behavior and Human Decision Processes, 42, 271-283.

Shaw, M.E. (1981). Group dynamics: The social psychology of small group behavior. New York: McGraw-Hill.

Simon, H.A. (1978). Information processing theory of human problem-solving. In W.K. Estes (Ed.), Handbook of learning and cognitive processes. Vol. 5. Hillsdale, NJ: Erlbaum Associates.

Slovic, P. (1966). Cue consistency and cue utilization in judgment. American Journal of Psychology, 79, 427-434.

Slovic, P. (1975). Choices between equally valued alternatives. Journal of Experimental Psychology: Human Perceptions and Performance, 1, 280-287.

Smart, C. & Vertinsky, I. (1977). Designs for crisis decision units. Administrative Science Quarterly, 22(4), 640-657.

Snadowsky, A. (1972). Communication network research: An examination of controversies. Human Relations, 25, 283-306.

Snyder, M., & Swann, W.B., Jr. (1978). Hypothesis-testing processes in social interaction. Journal of Personality and Social Psychology, 36, 1202-1212.

Stake, R. E. (1967). The Countenance of educational evaluation. Teachers College Record, 68(7), 523-540.

Stake, R.E. (1970). Objectives, priorities, and other judgment data. Review of Educational Research, 40(2), 181-212.

Stake, R.E. (1972). Focus or portrayal. Evaluation and Measurement Newsletter, 14, 1-2.

Stake, R.E. (1973). Measuring what learners learn. In E. R. House (Ed.) School Evaluation, The Politics and Process. Berkeley, CA: McCutchan Publishing.

Stake, R.E. (1975). Program evaluation, particularly responsive evaluation (Occasional Paper, No.5). Kalamazoo, MI: Western Michigan University Evaluation Center.

Stake, R.E. (1982). A peer response: A review of program evaluation in education. When? How? To what ends? In E.R. House, S. Mathison, J.A. Pearsol, & H. Preskill (Eds.), Evaluation studies review annual (vol 7). Beverly Hills, CA: Sage.

Staw, B.M. (1981). The escalation of commitment to a course of action. Academy of Management Review, 6, 577-587.

Staw, B.M., Sandelands, L.E., & Dutton, J.E. (1981). Threat-rigidity effects in organizational behavior: A multilevel analysis review. Administrative Science Quarterly, 26(4), 501-524.

Stolovich, H.D. (1975). Formative evaluation of instructional games. Improving Human Performance Journal, 4(3), 126-141.

Stolovich, H.D. (1976). How to produce audiovisual modules: Evaluating your product. NSPI Journal, 15(4), 3-5.

Stolovich, H.D. (1982). Applications of the intermediate technology of learner verification and revision (LVR) for adapting international instructional resources to meet local needs. NSPI Journal, 21(7), 16-22.

Stufflebeam, D.L. (1968). Toward a science of educational evaluation. Educational Technology, 8(14), 5-12.

Stufflebeam, D.L. (1983). The CIPP model for program evaluation. In G.F. Madaus, M. Scriven, & D.L. Stufflebeam (Eds.), Evaluation Models: Viewpoints on Educational and Human Services Evaluation. Boston, MA: Kluwer-Nijhoff.

Stufflebeam, D.L., Foley, W.J., Gephard, W.J., Guba, E.G., Hammond, R.L., Merriman, H.O., & Provus, M. M. (1971). Educational Evaluation and Decision-making. Itaska, IL: F. E. Peacock Publishers.

Taba, H., & Sawin, E. (1962). A proposed model in evaluation. Educational Leadership, 1962, 57-71.

Taylor, S.E. (1983). Adjustment to threatening events--a theory of cognitive adaptation. American Psychologist, 38(11), 1161-1173.

Tucker, S.A., & Dempsey, J.V. (1991). An evaluation semiotic. American Journal of Semiotics, 8(4), 73-103.

Tushman, M. (1979). Work characteristics and subunit communication structure: A contingency analysis. Administrative Science Quarterly, 24, 82-98.

Tversky, A. (1972). Elimination by aspects: A theory of choice. Psychology Review, 79, 281-299.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. Cognitive Psychology, 4, 207-232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science, 211(4481), 453-458.

Tyler, R.W. (1942). General statement on evaluation. Journal of Educational Research, 30, 492-501.

Wager, L.C. (1983). One to one and small group formative evaluation: An examination of two basic formative evaluation procedures. Performance and Improvement Journal, 22(5), 5-7.

Wallsten, T.S. (1968). Failure of predictions from subjectively expected utility theory in a Bayesian decision task. Organizational Behavior and Human Performance, 3, 239-252.

Walsh, J.P., & Fahey, L. (1986). The role of negotiated belief structures in strategy making. Journal of Management, 12(3), 325-338.

Walsh, J.P., Henderson, C., & Deighton, J. (1988). Negotiated belief structures and decision performance: An empirical investigation. Organizational Behavior and Human Decision Processes, 42, 194-216.

Weckworth, V.E. (1969). A conceptual model versus the real world of health care service delivery. In Hopkins, C.E. (Ed.) Outcomes Conference: Methodology of identifying,measuring and evaluating outcomes of Health Service Programs, Systems and Subsystems. Rockville, MD: Office of Scientific and Technical Information, NCH SRD, 39-62.

Weiss, H.M., & Knight, P.A. (1980). The utility of humility: self-esteem, information search, and problem solving efficiency. Organizational Behavior and Human Performance, 25(2), 216-223.

Wendt, D. (1969). Value of information for decisions. Journal of Mathematical Psychology, 6(3), 430-443.

Weston, C.B. (1986, Winter). Formative evaluation of instructional materials: An overview of approaches, Canadian Journal of Educational Communications , 15(1), 5-11.

Wyer, R.S., Jr., & Srull, T.K. (1980). The processing of social stimulus information: A conceptual integration. In R. Hastie, T.M. Ostrom, E.B. Ebbesen, R.S. Wyer, Jr.., D.L. Hamilton, & D.E. Carlson (Eds.), Person memory: The cognitive basis of social perception (pp. 227-300). Hillsdale, NJ: Erlbaum Associates.

Zemke, R. & Kramlinger, T. (1982). Figuring things out: A trainers guide to needs and task analysis. Reading, MA: Addison-Wesley.

Figure 1. Summary of criteria for three evaluation models: objectives-based (OB), decision-based (DB), and values-based (VB)

MODEL CRITERIA	OB	DB	VB
1. intended outcomes documented	•	•	•
2. unintended outcomes documented		x	•
3. document contexts leading to outcomes		•	•
4. document processes leading to outcomes		•	•
5. client standards overt		•	•
6. evaluator standards overt			•
7. client-evaluator negotiation			
8. program improvement-oriented			•
9. hard & soft data balanced			

key: • denotes criterion being met in a model

x denotes criterion being inconsistently met in model

Figure 2. Polemics of an Evaluation

Questions: To what degree should they be...

1. ( ) ( ) ( ) ( ) ( )
client helpful — field helpful

2. ( ) ( ) ( ) ( ) ( )
formative — summative

3. ( ) ( ) ( ) ( ) ( )
objectives-based — holistic-based

4. ( ) ( ) ( ) ( ) ( )
independent — negotiated

Sources: To what degree should they be...

5. ( ) ( ) ( ) ( ) ( )
descriptive — judgmental

6. ( ) ( ) ( ) ( ) ( )
quantitative — qualitative

7. ( ) ( ) ( ) ( ) ( )
in-dwelling — one-shot

8. ( ) ( ) ( ) ( ) ( )
deductive — abductive — inductive

Standards: To what degree should they be...

9. ( ) ( ) ( ) ( ) ( )
tacit — overt

10. ( ) ( ) ( ) ( ) ( )
proactive — reactive

11. ( ) ( ) ( ) ( ) ( )
self-initiated/internal — imposed/external

12. ( ) ( ) ( ) ( ) ( )
elitist or parochial — subordinate

Figure 3: 4 phase chart

1076

FIGURE 3

INPUTS WITHIN CONTEXTS	PROCESSES	OUTPUTS
PHASE 1: QUESTIONS		
INFORMATION	PERCEPTUAL REPRESENTATION	ACTION, RESPONSE, AND/OR JUDGMENT
1. Client's Task Demand • surface and depth factors • stable vs unstable contexts • past performance of clients • client belief structures	**2. Evaluator's Perceived Task** • prior experience of evaluator • semantic/semiotic content priority • confidence, politics, resources, audiences, purposes, boundaries	**3. Evaluator sets Questions and Standards** • phases: analysis, design, development, implementation
4. Client feedback Message to Evaluator • client-helpful • field-helpful • audience-helpful • expectations • spontaneous vs. systematic	**5. Evaluator Represents Feedback** • effort, rapport • match mismatch • objectives, management, or perception-based • changes between task & feedback: overt/covert • formative vs. summative • internal vs. external	**6. Negotiate Paradigm and Schema Setting** • negotiated vs. independent • response certitude, framing • self-set vs. assigned standards

INPUTS WITHIN CONTEXTS	PROCESSES	OUTPUTS

PHASE 2: DESCRIPTIONS

7. Evaluator collects data

- qualitative vs. quantitative
- stable vs. changing
- indwelling vs. 1-shot
- intuitive vs. analytical
- maximal/optimal/actual
- disjunctive vs. conjunctive
- additive vs. dicretionary

8. Evaluator analyzes data

- information load
- description vs. judgement
- anticipated vs. unanticipated
- analysis model
- audience analysis

9. Evaluator Reports Initial Descriptions

- feedback time: simultaneous vs equential prove/improve

10. Client Feedback Message to Evaluator

- initial certitude
- credibility of Evaluator
- match-mismatch
- selective perception

11. Evaluator Represents Client's Feedback

- prove/improve
- quantitative, qualitative or holistic
- overt vs. covert
- subjective, intersubjective, objective
- balance of positive/negative feedback

12. Evaluator Edits Descriptions and Reports Alternatives

- feedback time
- framing
- heuristics vs. algorithyms
- implications

Inputs Within Contexts	Processes	Outputs
Phase 3: Judgements		
Information	**Perceptual Representation**	**Action, Response, and/or Judgment**
13. Client's Feedback Message to Evaluator • expect gains or losses • truth vs utility • Client's certitiude • purposes & resources • selective perception • strategic myopia	**14. Evaluator's Editing of Alternatives** • concrete vs abstract • duration of events • display format • self efficacy	**15. Evaluator Rates Alternatives** • feedback processing time
	16. Evaluator/Client Judgment • inductive, deductive, abductive • heuristic vs normative; • confidence of Evaluator, • expectations: internal vs external • overt vs covert; proactive vs reactive • uncertainty of decision: cognitive dissonance	**17. Delivery Matrix & Recommendations** • time and format of delivery

Inputs Within Contexts	Processes	Outputs

Phase 4: Decisions & Choices

Information	Perceptual Representation	Action, Response, and/or Judgment
	18. Decision Mechanism • satisfice • self-initiated vs imposed • role of Evaluator and Client in decisions • risk taking vs risk seeking: error penalties • novice vs exper	
	19. Choice Mechanism • perceived self-efficacy • environmental demands (closed, open or optimizing system) • elitist vs subordiante • attractiveness of alternatives/ incentives • past expenditures of effort • performance valence: short & long term • press/compensate/integrate/ inaction	**20. Outcomes & Implementation**

Table 1. Criteria for Evaluating Instructional Hypermedia

1. Intertextuality and Intermediality

- what is the nature of the learner (prior knowledge, reflexive skills, expectations, motivation)?
- what is the instructional model used (e.g., eclectic or theory based)?
- what supports (e.g., focusing, hints and shaping) are provided for the user/learner ?
- what levels of information are included and what is the size and range of the knowledge base?
- how is information related (e.g., synchrony versus historical, diffuse versus focused, traditional work structures versus dynamic culture and discourse)?
- how is linear and nonlinear information linked?
- how is content granularity developed (e.g., how is relevant from irrelevant information filtered, chunking, degree of modularization)?
- how can hypermedia be structured to replicate content structures or knowledge structures?
- how is remediation versus enrichment information access provided?
- as instruction becomes richer, how does abstraction emerge (and is it backed by sufficient examples, practice, and alternative representations in different modalities)?

2. Decentering and Recentering

- how does the instructional model impose organization?
- how does the software impose organization: hierarchically and referentially?
- how are learner and teacher contexts defined and managed over time?
- what kind of representations are included in the courseware: hierarchical, relational and/or dialectical?
- what are the possibilities revealed by alternative sign systems?
- what objects are used in representations? how realistically are situations protrayed?
- what is the understanding of the conceptual structure of the information by learner and designer?
- what is the impact of different sequences of decentering and recentering?
- how do users assess different representations?
- how is decentering skill related to explicit organization and indivdual knowledge structures?

3. Navigating Networks: Achieving Cognitive Search Space

- how is exploration conceived by the developers? what charting procedures exist?
- how do learners use their increasing power over the sequencing of material to gain meaning?
- to what extent are hypermedia users (and developers) bound by acculturation to book technology?
- when and how are links denoted meaningfully (at the start of a node or within it)? Of a related nature, should specific parts of the screen be reserved for iinks? When should links be imposed and when should they be learner defined?
- how many nodes can be displayed at one time without being confusing?
- how do learners avoid getting lost in "hyperspace"? how much support exists for "dynamic" linking? (e.g., mapping and audit trails)
- how can designers accommodate to both self-learners and those needing more external structure?
- when is it effective to insert critical questions, navigational guidance or hints to users?
- how can higher order thinking like hypothesis formation be prompted?
- how can various imagery and sounds/intonations be used to access emotions?
- when do sound and visual realities need to be separated for user load given different symbolic systems being used?

4. Boundaries between Experts and Novices

- what methods for information retrieval are available?
- what methods are used for structuring hypermedia: deductive, inductive &/or abductive?
- what methods of browsing are available (e.g., single word/phrase search, Boolean logic, alphabetical index of node names, graphic maps of node relations)?
- how well does planning match execution of novice and expert access strategies?
- how representative are learning levels of development team? Additionally, to what degree are they captured in a mindset (e.g., prescriptive) ?
- how effective is transition from novice to expert stages (i.e., when to use advisors such as online and offline instructional aides, use of adjunct and guiding questions, heuristics, modelling--and combinations thereof)?
- how smoothly can the learner move between two representations?
- to what degree are metacognitive or self-learning skills overtly taught?
- to what extent is heuristic guidance content specific in hypermedia contexts? when should information be suppressed?
- what are the motivational effects of learner control as they transition through the hypermedia environment?
- when is it beneficial for the learner to discover various paths on their own rather than via the minimal path?

5. Boundaries of Individual Work

- to what extent do metaphors emerge that help us conceptualize this complexity?
- to what extent are designers of hypermedia tacitly influenced by a print-based mentality?
- how does hypermedia influence an author's cognitive load (e.g., capacity to make decisions about links, content and transitions)?
- when does the author and the user perceive the significance of the link?
- how should links be denoted that have the same referent given hypermedia's capacity to indicate the relational strength of each node?
- what grammar can be established that portrays information non-linearly?